



Deductive reasoning from uncertain conditionals

Guy Politzer, Gaëtan Bourmaud

► To cite this version:

Guy Politzer, Gaëtan Bourmaud. Deductive reasoning from uncertain conditionals. British Journal of Psychology, 2002, 93, pp.341-385. ijn_00000652

HAL Id: ijn_00000652

https://hal.science/ijn_00000652

Submitted on 3 Dec 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

British Journal of Psychology, 93, 354-381, (2002).

Deductive reasoning from uncertain conditionals

Guy Politzer¹ and Gaëtan Bourmaud²

¹C.N.R.S. , Saint-Denis, France

²University of Paris VIII, France

This paper begins with a review of the literature on plausible reasoning with deductive arguments containing a conditional premise. There is concurring evidence that people presented with valid conditional arguments such as Modus Ponens and Modus Tollens generally do not endorse the conclusion, but rather find it uncertain, in case (i) the plausibility of the major conditional premise is debatable, (ii) the major conditional premise is formulated in frequentist or probabilistic terms, or (iii) an additional premise introduces uncertainty about the major conditional premise. This third situation gives rise to non monotonic effects by a mechanism that can be characterised as follows: the reasoner is invited to doubt the major conditional premise by doubting the satisfaction of a tacit condition which is necessary for the consequent to occur. Three experiments are presented. The first two aim to generalise the latter result using various types of conditionals and the last shows that performance in conditional reasoning is significantly affected by the representation of the task. This latter point is discussed along with various other issues: we propose a pragmatic account of how the tacit conditions mentioned earlier are treated in plausible reasoning; the relationship of this account with the conditional probability view on conditional sentences is examined; an application of the same account to the Suppression Effect (Byrne, 1989) is proposed and compared with the counterexample availability explanation; and finally some suggestions on how uncertainty could be implemented in a mental logic system are presented.

People may have more or less confidence in the truth of the propositions that originate from their sources of information: communication, perception, memory, and inference. A person was told that P, saw that Q, recalls that R, infers that S; but was her informant dependable, her senses reliable, her memory faithful, her conclusion valid? People face such questions constantly (this is made possible by their metacognitive skills): they are more or less confident in the truth of the propositions they entertain. We take this fact as a psychologically primitive phenomenon and call degree of *belief* the subjective degree of confidence that people experience in such situations.

Belief comes by degrees: having full belief in a proposition is to consider it as true (hence full belief in not-P is to consider not-P as true, i. e. , P as false, which is full disbelief in P). One can have less than full belief in P, in which case one is uncertain (or doubts) about the truth of P; in particular, one can have slightly less than full belief, in which case one is slightly uncertain about P (and at the same time nearly certain about not-P): uncertainty about P is the extent to which one disbelieves P (within the limits where belief in P is greater than belief in not-P). When degree of belief decreases from full belief in P, a point of indeterminacy is reached where belief in P equals belief in not-P: This is the point of maximal uncertainty about P. The probability calculus assigns a probability equal to 1 to a true proposition and a probability equal to 0 to a false proposition. In one form of Bayesianism, the above mentioned point of maximal uncertainty receives a probability equal to 1/2. This choice is much debated and we need not take a stand on this issue because we need not assume that the standard probability calculus provides a psychologically appropriate format of representation of belief.

Belief is an individual's mental attitude in relation with a proposition at a point in time; it is helpful to talk of a level of *credibility* to characterise the proposition in question for that individual at that moment.

Plausible reasoning is pervasive in daily life as well as in scientific activity. While inductive reasoning and probabilistic thinking have been the object of much interest among psychologists for a long time, the frequent case where people process uncertain premises and draw an uncertain conclusion through a deductive argument has remained relatively neglected¹. This is so despite the recognition of its importance by logicians and mathematicians (Adams, 1975; Adams & Levine, 1975; Rescher, 1976; Suppes, 1966) and by philosophers (Pollock, 1987) and the

¹ It might be objected that a deductive argument cannot yield an uncertain conclusion. This is to forget the conditional definition of deduction: a deductively valid argument has a conclusion that is necessarily true *if its premises* are true. It follows that arguments that have been formally identified as deductively valid can receive uncertain premises (and yield an uncertain conclusion). The field so defined does belong to plausible reasoning while dealing with deductive arguments.

development of nonmonotonic reasoning formalisms in Artificial Intelligence. Research in this area is recent: most of the relevant work has been carried out in the past decade and, interestingly, it concerns nearly exclusively conditional arguments.

The main aim of this paper is to address the question of the conditions under which a deductive conditional argument yields an uncertain conclusion. An answer will be proposed based on an analysis of the literature; it will be followed by two experiments that aim to test the proposed answer while remedying some methodological shortcomings from previous work. Finally, an additional experiment will show the relevance of the results to standard conditional reasoning.

Review of the literature

In this section we will review a number of experimental manipulations carried out on deductively valid arguments that result either in a decrease in the endorsement of the conclusion, or in judgments of uncertainty about the conclusion (or both).

Varying the credibility of the premises

In an early investigation of causal conditionals, O'Brien, Costa, and Overton (1986) used a truth-table evaluation task, which indeed was a study of plausible reasoning although it was not presented as such. In the frame of medical vs mechanical scenarios, participants were presented with conditional sentences that expressed a hypothesis (e. g. *If the bone chips are removed, then the pain will be eliminated* or *If the thermostat is replaced, then the car will not overheat*). They were then given the result of an observation: it stated that an operation was performed [P] or not performed [not-P]) and the patient recovered [Q] (or did not recover [not-Q]) or that a part was replaced (or not replaced) and the engine did not overheat (or still overheated); all four combinations were proposed. Participants were then asked about the doctor's (or mechanic's) certainty that the hypothesis was correct (the options were: *certain that correct; certain that incorrect; cannot be certain*) in each of the four cases. One of the main results is that for the [P, Q] case (operation and recovery or part replaced and engine working), the hypothesis was more often estimated as uncertain in the medical scenario than in the mechanical one and that for the [P, not-Q] case, the hypothesis was less often estimated as falsified in the medical scenario (operation and no recovery) than in the mechanical one (part replaced and engine not working). According to the authors, these results are explainable by people's view of the medical domain, which is generally less deterministic than the mechanical one, so that medically there may be hidden internal causes that prevent an action being efficacious. In brief, this manipulation shows that information with the same logical

status can affect belief in a conditional sentence differently depending on the conceptual domain involved. The key factor seems to be the awareness that the level of understanding of the causal link between antecedent and consequent of the conditional differs from one domain to another: for the participants the link may be more or less strong, allowing for more exceptions to the hypothesis if it is weaker.

Cummins (1995; Cummins, Lubart, Alksnis, and Rist, 1991) developed a similar notion and made it operational. She analysed causal conditionals with respect to the number and the availability of factors that can prevent the effect from occurring, called disabling conditions. With Modus Ponens and Modus Tollens arguments² using causal conditional premises, the acceptance rating of the conclusion (on a 6-point scale ranging from *very sure that I cannot draw the conclusion* to *very sure that I can*) was a decreasing function of the number of disabling conditions. For example, of the following two MP arguments, *If the match was struck, then it lit; the match was struck / it lit*, and *If Joe cut his finger, then it bled; Joe cut his finger / it bled*, people accept less readily the conclusion of the first, which has many disabling conditions, than that of the second, which has few.

Thompson (1994, 1995) obtained similar results with causals, and also non causal rules such as obligations, permissions and definitions by using conditionals that varied in perceived sufficiency. She defined a sufficient relationship as one such that the consequent always happens when the antecedent does, whereas in a nonsufficient relationship the antecedent does not guarantee the consequent. The following are instances of a high level and of a low level of sufficiency, respectively: *If the licensing board grants them a license then a restaurant is allowed to sell liquor. If an athlete passes the drug test at the Olympics then the IOC can give them a medal*. The author observed that the rate of endorsement of the conclusion was an increasing function of the level of sufficiency (independently estimated by judges). Notice that, as can be seen in the two instances, the level of sufficiency depends on necessary conditions: with a low level of sufficiency many necessary conditions are missing whereas with a high level of sufficiency few necessary conditions are missing, as Cummins' analysis of disabling conditions would suggest (because the absence of a necessary factor for the effect to occur is logically equivalent to the presence of a disabling condition). This also applies to the study carried out by Liu, Lo, and Wu

² Throughout the present paper, reference will be made to the two basic deductively valid arguments used in studies of conditional reasoning, viz. (1) Modus Ponens (henceforth MP): If P then Q; P; therefore Q; and (2) Modus Tollens (henceforth MT): If P then Q; not-Q; therefore not-P. Two other arguments are also often studied, viz. (1) the fallacy of affirming the consequent: If P then Q; Q; therefore P; and (2) the fallacy of denying the antecedent: If P then Q; not-P; therefore not-Q. Because they are not deductively valid, they will not be considered in the present paper.

(1996). They considered three levels of "perceived sufficiency" for conditional statements. This variable was operationally defined by judges' perceived conditional probability of the consequent on the antecedent (a definition that could apply to Thompson's concept of perceived sufficiency as well). There were three levels: high (uncontroversial because definitionally true, e.g. *If John lives in Canada, then John lives in the northern hemisphere*), medium (expressing common regulations or habits, e.g. *If Mary cheats, then Mary will be punished by the school*) and low (expressing obviously debatable stereotypes, e.g. *If Bob wears glasses, then Bob is intelligent*). They observed that the rate of endorsement of the conclusion of MP and MT arguments was an increasing function of the level of perceived sufficiency.

George's (1995) studies also exploited conditional premises that are controversial or not fully credible. In his first experiment he used a two-step procedure. First, participants were asked to evaluate on a 7-point scale the credibility they attributed to causal conditional statements (e. g. *If the winter is harsh, then many people will be ill*). Second, the same statements were used in MP arguments whose conclusions had to be evaluated on the same scale. It was observed that there was a high correlation between belief in the conditional and belief in the conclusion, suggesting that the degree of belief in the conditional (the major premise of the argument) was conveyed to the conclusion. An even more direct evidence of the effect of belief in the conditional premise on the willingness to endorse the conclusion was provided by the third experiment. Two groups of participants received contrasted instructions. One group was asked to assume the truth of the conditional, whereas the other group was invited to consider its uncertainty. As a result, 60 percent in the first group endorsed the conclusion of at least three of the four MP arguments, whereas in the second group, only 25 percent did.

Introducing probabilistic expressions in the premises

George (1997, experiment 1) manipulated the degree of belief in the conditional premise of a MP argument by means of a probability expression modifying the consequent of the conditional. Two versions of the conditional were used as the major premise of MP arguments, e. g. *If Peter is in the kitchen, then it is very likely that Mary is in the garden* vs *If Peter is in the kitchen, then it is not very likely that Mary is in the garden*. The conclusion was evaluated on a 9-point scale. Nearly all the responses expressed some degree of uncertainty and the modal response was identical to the probability term used in the conditional, suggesting again the propagation of the degree of belief from premise to conclusion and the conservation of its value.

One can conclude from this group of studies that the conclusion of standard arguments such as MP and MT is rated as uncertain, and in accordance endorsed less often, when (1) either the conditional premise contains an explicit probability expression (usually modifying the consequent), which marks the degree of belief in the statement; (2) or the conditional is considered as doubtful because there are well-known factors that can affect the credibility of the consequent and there is uncertainty as to the presence or the absence of these factors. More precisely, the presence of some of these factors is sufficient to prevent the consequent from occurring, and the absence of some others is necessary for the consequent to occur. The first case above appears to be one in which the relationship is not specified enough to enable one to identify the possible factors of uncertainty.

Introducing additional premises

While the foregoing studies used standard MP and MT arguments, that is, arguments made of two premises, a major conditional and a minor, the next studies to be reviewed used a modified task of conditional reasoning that differed from the standard task by the presence of some information in addition to the pair of premises.

Byrne's (1989) article on what is now called the "suppression effect" offers a seminal study to which all subsequent research on reasoning with uncertain conditional premises has referred, although she did not present it as an investigation of uncertain reasoning. She had a first group of participants solve standard arguments such as, for MP: *If she has an essay to write then she will study late in the library; she has an essay to write; conclusion: (a) she will study late in the library; (b) she will not study late in the library; (c) she may or may not study late in the library.* A second group had to solve the same arguments modified by the addition of a conditional premise whose consequent repeated the consequent of the major and whose antecedent was another sufficient condition for the consequent to hold, such as *If she has some textbooks to read then she will study late in the library.* A third group had to solve the same arguments as the first group modified by the addition of a conditional premise whose consequent repeated the consequent of the major and whose antecedent was a necessary condition for the consequent to hold, such as *if the library stays open then she will study late in the library.* While, for the first two groups, there were above 95 percent of correct responses on MP and MT, for the third group this rate dropped to around 35 percent on both MP and MT, that is, the majority chose the uncertain conclusion (response (c)). In brief, when the antecedent of the additional conditional premise was another sufficient condition, it did not affect the rate of endorsement of the conclusion; but when it was a necessary condition, it

decreased the rate of endorsement of the conclusion. Byrne interpreted her results in terms of the suppression of valid inferences by contextual information. Notice that, at this stage, the fact that participants choose response (c) is compatible with two hypotheses. Either they process binary truth-valued propositions and their conclusion being neither true nor false they choose option (c) (as good reasoners do when they are asked to evaluate the conclusion of an invalid argument); or participants are engaged in plausible reasoning, in which case their response is a means (the only means in fact) of expressing their conclusion together with their degree of certainty about this conclusion. This might be revealed by a format of response with an appropriate scale of measurement. The next few studies will help to identify what we think is the correct hypothesis but notice first that in a recent series of experiments Byrne, Espino & Santamaría (1999) have confirmed the suppression effect with various refinements such as: (1) asking participants to produce their responses, (2) presenting expanded conditional premises in order to block the converse, (3) formulating the premises with *if and only if*, and (4) adding a fourth conditional premise that expresses an alternative sufficient antecedent. Some of these results have direct bearing on the issue just raised, but their examination, which is rather technical, is deferred to the general discussion.

Chan & Chua (1994) used various non causal conditional rules with MP and MT arguments. For each conditional premise, they defined three necessary conditions for the consequent of the conditional to hold ("additional requirements" in their terminology); these conditions varied in strength (that is, in degree of necessity or importance as estimated by judges). For example, with a MP whose major premise was *If Steven is invited then he will attend the party* the three levels of necessity were introduced each time by an additional premise following Byrne's (1989) paradigm: *If Steven knows the host well then he will attend the party* (or *If Steven knows at least some people well then he will attend the party*, or *If Steven completes the report to night then he will attend the party*). The response options were *he will attend the party*; *he will not attend the party*; *he may or may not attend the party*. It was observed that the rate of endorsement of the conclusion of these three-premise arguments was a decreasing function of the degree of necessity, and correlatively the rate of "*maybe or maybe not*" conclusions was an increasing function of the degree of necessity. In brief, the statement of an additional conditional premise that contained a necessary condition in its antecedent diminished the rate of endorsement of the conclusion all the more sharply as the necessity of the condition was high. This result sheds some light on the theoretical question raised above. If the evaluation of the conclusion was an all-or-nothing matter reflecting judgment of validity, it is hard to see how degrees of necessity could influence the response. On the other hand, if the non-endorsement

of the conclusion expresses degrees of belief in it, it is understandable that this belief varies with the degree of necessity of the additional condition. The next investigation provides more direct evidence in support of this view by attaching a measurement of uncertainty to the evaluation of the conclusion.

Stevenson and Over (1995, first experiment) modified Byrne's three-premise arguments by adding a fourth premise whose aim was to moderate the effect of the third. The first experimental condition was a standard argument, e. g. (for MP) *If John goes fishing, he will have a fish supper; John goes fishing; conclusion to be evaluated on a five-option scale: John will have a fish supper; . . will probably have. . ; . . may or may not have. . ; . .probably won't have. . ; . . won't have. . .* The second condition had a third premise whose antecedent was a necessary condition as used by Byrne: *If John catches a fish, he will have a fish supper.* The other five conditions had a fourth premise that informed the subject about the likelihood of the necessary condition's being satisfied: *John is always lucky; . ; . . almost always. . ; . . sometimes. . ; . . rarely. . ; . .very rarely.* While in the second condition Byrne's results were replicated, the effect of the fourth premise on both MP and MT was to decrease the rate of endorsement of the conclusion and to increase the expression of uncertainty on the five-point scale in a near-monotonic fashion across conditions. This clearly establishes that the manipulation of degrees of necessity results in functionally related degrees of belief in the conclusion.

In their second and third experiments, Stevenson and Over (1995) used three-premise arguments in which the second premise was a categorical sentence that introduced various levels of frequency directly into the necessary condition. For example, given the major premise *If John goes fishing, he will have a fish supper,* there were five levels in the second premise: *John always catches a fish when he goes fishing; . . almost always. . ; . . sometimes. . ; . . almost never. . ; . . never. . .* In the second experiment, the conclusion was evaluated on a five-option scale and in the third experiment the likelihood of the conclusion was evaluated on an 11-point scale. For both MP and MT the rate of endorsement of the conclusion decreased and the uncertainty scores (derived from the options chosen on the scale) increased monotonically as the frequency mentioned in the second (categorical) premise decreased (with a floor effect on the two smallest frequencies). The comparison of the results between the first two experiments is of special interest because participants were asked to assume that the premises were true in the first case but not in the second case. On all comparable conditions the rate of endorsement of the conclusion was much higher and correlatively the uncertainty scores much lower in the first experiment than in the second one.

Stevenson and Over concluded that "the impact of the additional premise in

Byrne's study was to induce uncertainty in the first premise" (p. 626). This means that by manipulating the degree of necessity one actually modifies the degree of belief in the major premise. This is so because the more strongly a necessary condition is satisfied, the stronger the associated sufficient condition is. From this perspective, Chan & Chua's use of various necessary conditions to manipulate levels of necessity is tantamount to introducing levels of belief in the major premise.

Manktelow and Fairley (2000) manipulated the extent to which a necessary condition is satisfied in such a way that a low degree operated as a disabling condition (when it is satisfied the consequent is less likely to occur) and a high degree operated as an additional requirement (when it is satisfied the consequent is more likely to occur). A standard MP argument with the major premise *If you pass your exams, you will get a good job* served as a control whereas the experimental groups received this MP with one of the following additional premises: (1) got very low grade; (2) got low grade; (3) got respectable grade; (4) got excellent grade. The conclusion had to be assessed on a 7-point scale (from very low to very high certainty to be offered a good job). For the first two groups the certainty ratings were below the control group (and lower for the *very low grade* group than for the *low grade* group). For the last two groups the certainty ratings were above the control group (and higher for the *excellent grade* group than for the *respectable grade* group). Again the degree of certainty of the conclusion is an increasing function of the degree to which a necessary condition is satisfied.

Hilton, Jaspars and Clarke (1990) used a different paradigm in which participants' interpretation of the conditional was inferred from their responses to the arguments. These authors presented their subjects with three sets of arguments. The first set consisted of a number of instances of the four standard arguments. An example for MP was: *If he works hard then he will pass; he works hard / he will pass*. To constitute the second set, these arguments were modified by the introduction of an additional categorical premise that affirmed an alternative cause for the consequent to hold: *If he works hard then he will pass; the exam is easy; he works hard / he will pass*. Similarly in the third set of arguments there was an additional categorical premise that denied the alternative cause: *If he works hard then he will pass; the exam is difficult; he works hard / he will pass*. For each set of arguments, on the basis of the responses (the conclusion was evaluated as: *True; sometimes true and sometimes false; false*) it is possible to infer participants' interpretation of the conditional premise. The authors classified these interpretations as expressing (1) sufficient (but not necessary) conditions; (2) sufficient and necessary conditions; (3) necessary (but not sufficient) conditions. Taking the standard two-premise arguments as a basis of comparison, arguments in which an alternative cause was asserted gave rise to more "sufficient"

interpretations of the conditional (and to fewer "necessary" interpretations); and arguments in which an alternative cause was denied gave rise to fewer "sufficient" interpretations of the conditional (and to more "necessary" interpretations). In brief, the logical characterisation of the interpretation of conditionals in terms of sufficiency (and necessity) was found to be sensitive to the presence of additional premises that affirmed or denied an alternative causal condition to the antecedent, with more sufficiency interpretations in the former case and fewer in the latter.

Using performatives

Newstead, Ellis, Evans, and Dennis (1997) have reported differences in the rate of endorsement of the conclusion as a function of the content of the conditional; in particular, promises and threats on the one hand, and tips and warnings on the other hand seem to constitute two contrasted groups, the former giving rise to more frequent endorsements of the conclusion than the latter (a result confirmed by Evans and Twyman-Musgrove, 1998). As noted by the authors, the key factor seems to be the extent to which the speaker has control over the occurrence of the consequent, which is higher for promises and threats than for tips and warnings. This control in turn determines the strength of the link between antecedent and consequent, a concept akin to that of degree of sufficiency. The authors claim that the type of truth table inferred from participants' performance is explainable in terms of uncertainty, a conclusion foreshadowed by the Hilton et al. study; they also claim that this is so because some conditionals lead to more believable contingencies than do others, another conclusion anticipated by the O'Brien et al. study: a remarkable coherence appears when these studies are brought together.

Belief revision paradigm

Finally, there is a newly developed paradigm in psychology initiated by research in Artificial Intelligence which is highly relevant to the present topic, namely belief revision (Elio and Pelletier, 1997). It deals with how people accommodate a new piece of information that contradicts some of their previous beliefs. If that information is taken as certain, people have to alter the set of beliefs that is concerned, by questioning or denying one or several propositions hitherto accepted as true. A first point of interest resulting from these studies is that a strong majority of participants opt for an expression of doubt rather than for a categorical denial of the proposition to revise (Dieussaert, Schaeken, De Neys, & d'Ydewalle, 2000; Politzer & Carles, 2001), confirming the pervasiveness of plausible reasoning.

Elio (1997) presented participants with MP and MT whole arguments, that is, premises (assumed to be true) followed by the categorical statement of the

conclusion: *If Joe cut his finger then it bled; Joe cut his finger; therefore you believe his finger bled*, after which the conclusion was declared to be false: *You discover that Joe's finger did not bleed*. The task was for participants to indicate on a 7-point scale what was then their degree of belief in each premise. When the conditional premise was a causal with few disabling factors, participants gave a higher degree of belief than when it had many disabling factors. This is direct evidence that necessary conditions are a factor of belief in conditionals. This conclusion is further supported by the results of Elio (1998) who, with the same paradigm (that is, a full conditional argument followed by the denial of the conclusion) asked participants to indicate how they would modify the premise they disbelieved as a result of the denial of the conclusion in order to restore consistency. Most of the responses consisted in either qualifying the conditional by defaults ("usually if P then Q"; "If P then Q but there are exceptions", etc.) or introducing a necessary condition conjoined with the antecedent of the conditional, like in *If Joe cut his finger and the cut was deep enough then it bled*, showing again how the satisfaction of necessary conditions controls the credibility of the conditional.

It is time to extract the common core of the foregoing studies. Each one in its own way conceptualises an apparently different variable (e.g. , number of disabling conditions, level of perceived sufficiency, presence of an alternative cause, strength of necessary conditions, frequency of necessary conditions) the result of which is to modify the degree of belief in the conclusion of MP and MT arguments and to affect accordingly their rate of endorsement. In fact, underlying those variables, there is one single common mechanism, namely the recognition of one or several factors that are necessary conditions for the consequent to occur and which, by this very fact, are conditions that implicitly complement the antecedent of the conditional to make it an actual sufficient condition. The degree of belief in the satisfaction of those factors acts as a mediator to define the degree of sufficiency of the conditional premise, that is, its credibility, and consequently, by inheritance, the degree of belief in the conclusion of the argument. The truth status of the conclusion is treated by degree rather than in an all-or-nothing manner and this degree is closely correlated to the degree of belief in the premise.

Consider, for instance, *If the match was struck, then it lit*. That the match be dry is one of the necessary conditions for the match to light, a condition whose satisfaction is not totally certain. That the match was struck is no longer a sufficient condition for the match to light if there is a doubt in the match's dryness, which is why upon hearing about the dryness of the match one may withhold full belief in the conditional. In order to restore full belief in it, the antecedent would have to be complemented with the necessary condition, *the match was dry*, yielding: *If the match*

was struck and it was dry, then it lit, which is why such factors can be called complementary necessary conditions (henceforth CNC). In case there is a doubt on this condition, the conclusion of the MP, *it lit*, is uncertain and inherits the degree of belief in the dryness of the match, whereas the conclusion of the MT, *the match was not struck*, knowing that it did not light, is also uncertain to the same degree. The concepts used by various authors can be unified. In the example chosen, dryness is the relevant CNC. Dampness would be a disabling condition in Cummins' terms; the levels of perceived sufficiency of the conditional are defined by the likelihood that either the CNC under consideration is satisfied (the match is dry) or a number of such factors are satisfied. Dryness is Hilton et al's "alternative cause"³ and the frequency of the cases where matches are dry is the variable considered by Stevenson & Over; finally, Chan & Chua's strength of necessity refers to the necessary aspect of the CNCs and to the fact that some are more important than others for social, psychological, physical reasons, etc. (dryness may be more important than absence of wind).

Even though the results of the studies that have been reviewed are remarkably coherent, some methodological improvements seem desirable. None of these studies respected the following features jointly: (1) using a standard condition that serves as a control; (2) defining treatments in such a way that the credibility of the conditional could not only be decreased but also increased; (3) considering more than two degrees of credibility; (4) keeping the major and minor premises and/or the context constant across levels of credibility; (5) using a format of evaluation of the conclusion sensitive enough to enable the expression of various degrees of belief; (6) varying the type of conditional sentences.

In the first experiment reported below, an attempt was made to satisfy all these methodological demands:

- with reference to (1), (2), and (3) above, there was a control condition (a standard argument), and three degrees of credibility with both increased and diminished credibility for every context kept unchanged.

- (4) for a given conditional sentence, a decrease in credibility was determined

³ In the examination example, *If he works hard then he will pass*, there is a confounding between the negation of a disabling condition and an alternative cause: the easiness of the exam is both the negation of a disabling condition (a difficult exam) and an alternative cause (i. e. another sufficient condition to pass the exam). This causal structure differs from the match example where the dryness of the match is the negation of a disabling condition but is not an alternative cause (unless dry matches light spontaneously). This is linked to the fact that *dry* and *damp* are contradictory, *dry* being the normal state of the world, whereas *easy* and *difficult* are contrary, with the neutral point as the normal state. The confounding shows that "alternative cause" in the Hilton et al. study is a misnomer.

by introducing an additional premise referring to the non-satisfaction of necessary conditions varying in importance, and an increase in credibility by introducing an additional premise referring to the satisfaction of a necessary condition, so defining a scale of credibility for the conditional from low to high. In addition, frequency or probability terms were generally avoided in defining the various levels of credibility, lest the task be too transparent. It was hypothesised that the ordered degrees of belief in the conditional operationalised by the mention of various degrees of satisfaction of Complementary Necessary Conditions would propagate to the conclusion of a conditional argument such as Modus Tollens.

- (5) the response format was a five-point scale of certainty.

- (6) four different types of conditionals were used: causal, means-end, decision, and remedial rule. This latter point, like the previous ones, meets a methodological precaution: it was thought that it was better to use a variety of types (even if limited to four) rather than only one type (e. g. causals) in order to avoid domain specificity. These categories are defined in relation to people's action, they are clearly pre-theoretical and it was not the aim of this experiment to make differential predictions between them.

EXPERIMENT 1

Method

Materials

Participants were presented with arguments made of two or three premises (one of which was a conditional sentence) and a conclusion to evaluate on a five-point scale (*true; probably true; indeterminable; probably false; false*)⁴. Five different contexts were created. For each of them, the characters and the situation were introduced in one sentence and were immediately followed by the premises and the conclusion to evaluate. The introductory sentence also served to inform that the antecedent of the conditional premise was a general and regular event.⁵ The minor premise and the

⁴The word "indeterminable" that is used here to denote the midpoint of the scale is the translation of French "*incertaine*", qualifying the conclusion. This latter word, unlike English "uncertain", could not refer to the participant's own uncertainty ("I don't know") for semantic reasons, and more importantly, for grammatical reasons: the adjective takes the feminine gender, in agreement with the word "*conclusion*" which it unambiguously qualifies.

⁵Setting a high frequency base rate for the antecedent was necessary for measurement reasons. This was to avoid that participants assume that the antecedent has a low frequency base rate, which (as a pilot study indicated) would invite them to avoid the low certainty end of the scale when assessing the negation of the antecedent (the conclusion of the argument).

conclusion referred to a special occasion where the major premise applied. One example of a context is given in full in the appendix.

The conditional sentences that appeared in the arguments (presented in the appendix) can be characterised as: one remedial rule (Headache), two means-end rules (Baker's and Fishing), one causal rule (Alarm), and one decision rule (Traffic). The following description applies to all five contexts.

A standard Modus Tollens argument with two premises and a conclusion was used to define the *standard condition*. This was intended to provide a base-line. There were four other conditions characterised by the addition of a third premise, the *additional premise*. Three of the conditions differed by the degree of belief in the major premise (the conditional sentence) that the additional premise induced. There were: a *low credibility condition* defined by an additional premise that stated that a condition necessary for the consequent to hold was *not* satisfied, and a *very low credibility condition* defined by an additional premise that stated that a condition *strictly* necessary for the consequent to hold was *not* satisfied⁶. The definition of these two levels results immediately from the theoretical considerations on the role of the CNCs. In line with the same considerations, suppose that one of the basic requirements suggested by world knowledge for the consequent to occur is known to be satisfied: then, it follows that the credibility of the conditional is at least as high as in the absence of such information. For, obviously it cannot be lower; and it can be higher to the extent that, should the hearer have any doubt as to the satisfaction of the various CNCs, at least the one that is mentioned is explicitly guaranteed. Accordingly, a *high plausibility condition* was defined by an additional premise that stated that an important condition necessary for the consequent to hold was satisfied.

Finally, there was a fifth condition (the Explicit-normality condition), in fact a pseudo-condition; it was always presented last, so that it could not have any effect on the responses to the other four conditions. Its objective was to know the effect of an additional premise, which made the normality assumption explicit by stating that "all the conditions for [the consequent to hold] were satisfied". Two opposite hypotheses could be made. One, the effect of such an additional premise would be to render the conditional highly credible and therefore to turn this condition into a very high credibility condition. But the other hypothesis was that by making explicit such an assumption that is normally tacit, the communicator invites the reasoner to infer

⁶ The difference in necessity of the additional statements used in the Low and Very Low conditions, which seems intuitively uncontroversial, was confirmed by participants in two pilot studies who served as judges.

that, after all, there might be yet other unknown conditions to be satisfied. In that case, the additional statement might have a counter-productive effect and fail to increase, or even diminish, the credibility of the conditional.

The arguments were presented in booklets. Each booklet contained instructions on the first page followed by five arguments, each on a separate page, in such a way that all five conditions and all five contexts appeared in a given booklet. This defined a number of combinations of scenarios by conditions equal to $5! = 120$, hence a number of 120 booklets (and as many participants).

Two orders of presentation of the conditions were chosen. One was: (1) standard condition; this first position provided a kind of calibration of the population; (2) low credibility; (3) high credibility; (4) very low credibility; (5) explicit-normality condition. The other order differed just by the exchange of the second and the fourth conditions. All possible orders of presentation of the contexts were used with equally balanced frequencies. The five sets of arguments are presented in the appendix.

Design

There were five contexts and five levels of additional premise, viz. one standard condition (absence of additional premise); three levels of credibility induced by the additional premise: High; Low; Very low; one explicit-normality condition. Levels of additional premise and contexts were within-subjects factors.

Participants and procedure

One hundred and twenty applicants for one of the french Air Force Officer's Academy participated by groups of twenty. All held at least a Secondary School Certificate with Mathematics and Science as a major subject and most of them had studied those subjects for two years at a Tertiary institution. However, they had no training in formal logic. The experiment took place at the same time as the admission interviews, which warranted a high level of motivation. The task was presented as a reasoning task. Each participant was randomly attributed a booklet on which he wrote his answers in the presence of the experimenter, working at his own pace.

Predictions

It was predicted that: (1) for the three manipulated conditions, the degree of belief in the conclusion would be an increasing function of the credibility of the conditional; in other words, the following order of belief in the *conclusion* should be observed: High credibility condition > Low credibility condition > Very low credibility condition; (2) the standard condition being no higher than the High credibility

condition and higher than the Low and Very low credibility conditions, the degree of belief in the conclusion of the standard condition should be situated accordingly, that is: High credibility condition \geq Standard condition $>$ Low and Very low credibility conditions; hence the final prediction for the order of belief in the conclusion:

High credibility condition \geq Standard condition $>$ Low credibility condition $>$ Very low credibility condition.

To the extent that the foregoing inequalities are expected to be general, they should apply separately and individually to each context.

Results

Table 1 presents the distribution of the responses (in percent) on the categories of evaluation of the conclusion for each context separately, and also after pooling across contexts, as a function of the credibility of the conditional premise.

Insert Table 1 about here

We begin with the analysis of the first four rows of the sub-tables, leaving aside for the moment the explicit-normality condition. According to the hypothesis, the expression of belief in the conclusion (measured on an ordinal scale) should be an increasing function of the credibility of the conditional (also measured on an ordinal scale). The link between these two variables can be appropriately evaluated by using Kendall's tau coefficient of correlation for ordered contingency tables. This was performed for each context separately. The value of tau was found to lie in the range of .48 to .71 (Headache: $\tau = .69$; Baker's: $\tau = .63$; Fishing: $\tau = .48$; Alarm: $\tau = .66$; Traffic: $\tau = .71$). A test of significance for these correlations was performed using the Jonckheere test for trends. All five values were found to be very highly significant (all $\underline{z}_c > 3.90$, $p < 10^{-5}$ and beyond). In other words, the present results support the hypothesis that the more credible the conditional premise (as measured by the four degrees of satisfaction of Complementary Necessary Conditions that have been defined) the greater the degree of belief in the conclusion of the argument (as measured by the five levels in the scale). This occurred on each context separately.

Another more refined, but also more local, way of testing the prediction consists in focusing on the first and last columns of the tables: there should be a decrease in the frequency of True responses and an increase in the frequency of False responses across levels of credibility, from High to very Low. (Note that this applies to the two extreme rows of the tables only; there is no reason to expect a monotonic

increase or decrease for the three internal columns). The Jonckheere test for trends was used. For the True response it yielded significant values ranging from $z_c = 2,98$ ($p < .01$ for the Fishing context) to $z_c = 6,34$ ($p < 10^{-6}$ for the Alarm context). For the False response, the results were also significant (ranging from $z_c = 2,57$, $p < .01$ for the Baker's context to $z_c = 4,70$, $p < 10^{-6}$ for the Traffic context) with the single exception of the Fishing context ($z_c = 1,54$, $p > .05$). This means that, in nine of the ten cases, there was a reliable overall increase (or decrease) in frequency of True (or False) responses across levels of plausibility, as expected.

Regarding the explicit-normality condition, on all contexts except the Traffic context the comparison of the distribution with the High credibility condition shows that the values are often very close. The Jonckheere test for trend applied to each of the two-row tables defined by these two conditions was calculated. It never approached the .05 level of significance, except on the Traffic context where a highly significant difference was found, showing that on this context the Explicit-normality condition yielded a distribution paradoxically as extreme as the Very low condition⁷.

A last point of interest concerns the *indeterminable* column. On all five contexts, it exhibits a remarkably clear inverse U-shaped trend across the four levels of premise credibility, the peak occurring on the Low level. This is exactly what should be observed when one plots uncertainty against degrees of belief (the minimal uncertainty coinciding with total belief or total disbelief, and the maximal uncertainty with an intermediate degree of belief). This verifies both the between-subjects consistency and the validity of the scale (the validity of the hierarchy for credibility levels being warranted on independent grounds).

Discussion

Before drawing any conclusion, a possible source of misunderstanding in relation to the use of different scenarios has to be dispelled. It was not the aim of this study to test between contexts; no hypothesis was made with respect to possible differences

⁷ This is what should be observed if participants interpreted the conditions in the additional premise, "*all the conditions for her to take her car were satisfied*" as referring to the smoothness of the traffic (the antecedent of the conditional and not to other tacit conditions) from which it follows that they should deny that the traffic is not smooth. Why could such an interpretation occur? The following explanation, admittedly post hoc, is proposed. The Traffic context is the only decision rule, one where the term "conditions" may refer in priority to the antecedent of the conditional (the traffic is light), the satisfaction of which produces the decision to take action (Mary takes her car). In brief, the discrepancy seems due to an interaction between the wording of the instructions and the decision rule. Given that the correlation observed for this rule supports the main hypothesis as well as the other rules (even better in fact) this discrepancy is certainly insufficient to refute the claim that the results generalise beyond the causal rule.

as a function of context, and consequently similarities are more important than differences for our current purposes. These may nevertheless be considered to the extent that they may be suggestive for further work. Such variations do exist, but they are never in the sense of a departure from correlation which, as we have seen, is highly reliable. They concern essentially two phenomena. The first concerns the mode of the very Low condition whose position fluctuates across the last three columns (whereas the mode of the other conditions is located either within one single column or two columns). The other phenomenon is the fluctuation of the base-line: the rate of endorsement of the conclusion (response *True*) in the standard condition varies from 25 % (Traffic) to 58 % (Alarm and Baker's) with a mean of 43 %. It reflects a well known "content effect" in conditional reasoning and it is understandable that the percentages in the other cells differ across tables accordingly. Notice that the concept of CNC has the potential to explain the content effects: the base-line is more or less high depending on the credibility of the conditional, which in turn depends on the availability of CNS's, but once again, such a manipulation was not the aim of the present experiment. This question was indirectly addressed in the third experiment below⁸. It would be illusory to draw conclusions from the comparison of the tables in the absence of (1) theoretically-driven predictions, and (2) independently-gathered information regarding the relative credibility of the conditionals. Of course, in saying so, we both acknowledge a limitation of this experiment and identify an objective for future research, namely the need for both varying context and controlling for the credibility of the sentences. In addition, more contexts will need to be considered, in particular those that are not linked with people's actions.

With these reservations in mind, the main result is the confirmation that the degree of belief in the conclusion of Modus Tollens arguments, measured on a five-point scale, is an increasing function of the credibility of the conditional sentence operationalised by a four-level hierarchy of Complementary Necessary Conditions. This result was observed with five contexts defined by four types of conditional rules (causal, remedial, decision, means-end). It confirms and generalises previous

⁸ These two observations are linked to the fact that we have not defined an absolute scale of credibility for conditionals. The four levels are ordered with respect to the CNCs that have been used for each conditional in isolation. This means that if such a thing as an absolute scale could be defined, it might be the case that the absolute value of the very Low level on one scenario coincides with that of, say, the Low level on another scenario. The point is that there are too many unknown and uncontrolled factors to define the credibility of a conditional in an experimental setting absolutely (and therefore to manipulate it at will). What CNCs are normally considered as satisfied in a given population, what is the perceived probability of the antecedent (to be denied in the conclusion of MT), what personal experience participants have with the content are but a few factors that may affect the rates of response across contexts.

experimental results obtained on the basis of various theoretical approaches and helps interpret them in a parsimonious way based on a single conceptualisation: the antecedent of a conditional sentence is seldom a sufficient condition by itself; rather, knowledge of the domain suggests complementary conditions that have the status of necessary conditions but are normally tacitly assumed to obtain by the speaker. A conditional sentence is credible to the extent that the satisfaction of the CNC is warranted. Whenever a premise added to the conditional sentence introduces doubt on the satisfaction of a CNC, one enters plausible reasoning, that is, the conditional becomes more or less credible and the conclusion inherits a degree of belief that depends on that of the conditional.

A careful examination of the tables reveals one striking result: overall, the High credibility condition differed only marginally from the Standard condition. The Jonckheere test for trend applied to the two-row tables obtained from these two conditions was significant on the Traffic context only ($z_c = 2.01$, $p < .05$) and failed to reach the .05 level on the other contexts. In brief, the conditions whose departure from the standard condition contributed to the functional relation between the credibility of the conditional and the belief in the conclusion are mainly the Very low and Low conditions. Whereas it seems relatively easy to manipulate the uncertainty of the conditional by decreasing its credibility, it would seem more difficult to manipulate it by increasing its credibility, as if the standard condition represents a ceiling level. This conclusion is reinforced by the observation that the explicit-normality condition too yielded distributions that were very close to the standard condition. Of course, this latter observation could receive the pragmatic explanation outlined above (with the description of the fifth condition). But at this stage it seems more parsimonious to hypothesise that for both the High credibility and the Explicit-normality conditions another common pragmatic phenomenon occurred. As suggested above, conditionals are typically expressed under a *ceteris paribus* assumption to the effect that the CNCs are satisfied, an assumption that it is the role of the additional premise to question. Consequently, an increase in the credibility of the conditional (with respect to its standard formulation) by stating complementary necessary conditions would be hard, if not impossible, to obtain because such necessary conditions are already implicitly assumed to hold. There would be some kind of principle of idempotence at work: there is no effect in repeating a condition already assumed. There is, however, another hypothesis that cannot be excluded. In the first experiment, the High condition was always presented in the third position (and the Explicit-normality condition in the last position). Before they saw the High condition, participants always solved a Low or a Very Low condition argument in which they were exposed to contexts where some necessary condition was

unsatisfied, hence an effect of order: participants might have been prompted to assume spontaneously that a necessary condition was unsatisfied. This would have led them to give a lower rating for the High condition than if they had received this condition in the first position. This possibility was tested in the second experiment in which participants had to solve two arguments in the High plausibility condition, one before and one after solving a Low plausibility argument.

EXPERIMENT 2

Method

Materials

The arguments to solve were a subset of those used in the first experiment, namely, Modus Tollens arguments in standard form, and in the High and Low plausibility conditions (the Very low plausibility condition was left out). Four out of the five contexts were used (the Traffic context was left out). The task was presented in booklets that contained, after a page of instructions, four Modus Tollens arguments to solve as follows: (1) a standard (two-premise) argument; (2) a first three-premise argument in the High credibility condition; (3) a three-premise argument in the Low credibility condition; (4) a second three-premise argument in the High credibility condition. Each context appeared once in each booklet; all combinations of condition and context were used, so that all contexts appeared an equal number of times in each of the four arguments. In particular, each ordered pair of contexts used for the two High credibility conditions always had its counterpart in the reverse order. These constraints led to the constitution of 24 different kinds of booklets.

Design

There were two orders of presentation of the High credibility argument (before or after the Low credibility argument) and three levels of credibility that were within-subjects factors.

Participants and procedure

Sixty-five first year students of Mathematics at the University of Paris VIII divided into four groups of approximately equal numbers served as subjects. They answered at their own pace at the beginning of a class. About one third declared to have studied some elementary logic. However, as far as the comparisons of interest are concerned, the data showed no difference between this sub-group and the other students, so that the results of all participants were pooled.

Results and discussion

Although this experiment was not planned for purpose of replication, an examination of the distribution of the responses (pooled across contexts) for the three levels of plausibility that have been used shows the same general trend as in the first experiment (see Table 2).

Insert Table 2 about here

To answer the question addressed by this experiment, credibility judgements of the conclusion of the first and second High credibility arguments were compared for each participant. Judgements were at the same level on the scale in 25 cases; the second judgement was more frequently higher (23 cases) than lower (17 cases), a result in the opposite direction than hypothesised. (No *post hoc* attempt to explain this opposite difference is needed because it is not significant: Sign test, $p > .13$).

Conclusions of High credibility arguments did not receive lower credibility evaluations when they were solved after a Low credibility argument than before. This result eliminates the possibility of an effect of the Low credibility argument on the evaluation of the subsequent High credibility argument. It helps clarify the interpretation of the first experiment, indicating that the failure to increase belief in the conclusion by the presentation of satisfied conditions is a genuine effect and not an experimental artefact. In all likelihood, degrees of belief in the conclusion did not increase in the High credibility and the Explicit-normality conditions because participants already implicitly assumed the content of the additional premise. It is interesting to compare this conclusion to data obtained by Manktelow & Fairley (2000). In their experiment described above it was observed that the presentation of 'additional requirements' (i. e. enabling conditions, *got respectable grade*, and especially the 'exaggerated form', *got excellent grade*) resulted in certainty ratings higher than the baseline defined by the standard MP. However, this is not incompatible with the present results because the content of these additional premises cannot be implicitly assumed in the standard premises precisely because it is situated above the normal conditions (which are something like '*got average grade*'); whereas if, as was the case in the first experiment, the content of the additional premise coincides with the normal conditions (e. g. normally all the conditions for the alarm to be set off should be satisfied) this content is already assumed and no increase in certainty is possible by a ceiling effect.

The second experiment studied the question of the possible effect of solving a Low credibility argument on a High credibility argument and answered in the

negative. However, the same question arises as far as *standard* arguments are concerned: could they be influenced by the solving of a previous argument? That is, participants might evaluate the conclusion of a standard Modus Tollens as less credible after solving Low credibility arguments than before. If this were the case, it would mean that not only is the deduction influenced by the consideration of the satisfaction of Complementary Necessary Conditions for the antecedent to hold, but more subtly, it also depends on an implicit license to consider and use such conditions, or in other words, on the representation of the task. This was studied in the third experiment.

EXPERIMENT 3

Method

Materials

The materials were a subset of those used in the first experiment. Four out of the five contexts were used (the Traffic context was left out) and three conditions were considered (Standard, Low credibility and Very Low credibility). Participants were presented with booklets that contained four Modus Tollens arguments to solve. All booklets contained, in the following order: (1) a first standard two-premise argument; (2) a three-premise argument in the Low credibility condition of the first experiment; (3) a three-premise argument in the Very Low credibility condition of the first experiment; (4) a second standard two-premise argument. Each context appeared once in each booklet; all combinations of condition and context were used, hence 24 different booklets.

Participants and procedure

One hundred and twenty applicants were sampled from the same population as the one used in the first experiment (in which none of them had participated). The procedure was identical.

Design

There were two positions in the sequence of arguments for the Standard argument (first or fourth), three levels of credibility of the conditional premise, viz. Standard, Low and Very low, and four contexts. Position of arguments, levels of credibility and contexts were within-subject factors.

Results

Table 3 presents for each context separately, and also after pooling across contexts,

the distribution of the responses (in percent) on the categories of evaluation of the conclusion as a function of the credibility of the conditional premise.

Insert Table 3 about here

Notice first the remarkable similarity of these distributions with the distributions obtained in the first experiment for each context, respectively, and also for the collapsed data; in the latter case, most cells differ by only a few percent from their counterpart in the first experiment: the main data of the first experiment seem very robust.

Next, the comparison of the two Standard distributions shows a cross-over on all four contexts, with fewer "True" evaluations (about 18 % less) when the argument was presented in the last position than when it was presented in the first position. The numbers did not allow enough statistical power to test each context separately. Pooling the results across all subjects as shown in Table 4, it appears that out of the

Insert Table 4 about here

63 participants who rated the conclusion of the first argument as true, 30 (47.6 %) did not rate the conclusion of the second as true; whereas out of the 41 participants who rated the conclusion of the second argument as true, only 8 (19.5 %) had not rated the conclusion of the first as true. This difference is very significant (McNemar test, chi-square = 11.6, $p < 10^{-4}$). In brief, participants endorsed the standard Modus Tollens less often when it was presented in the last position than in the first position, that is, less often after answering two arguments that belonged to the Low and Very low credibility conditions in which the credibility of the conditional sentence was explicitly manipulated.

Discussion

It was hypothesised that people might evaluate the conclusion of a standard MT argument differently before solving modified arguments rather than after. In the two modified arguments, a third premise stated that a condition necessary for the major premise to be sufficient was not satisfied, hence the low degree of belief attributed to the conclusion of these two arguments. The hypothesis was that a transfer effect might occur from modified to standard argument: participants might evaluate the conclusion of a standard argument as less credible after solving the modified arguments. This is what actually happened. The rate of full belief in the conclusion decreased reliably by about 18%. In other words, the sheer solving of the

two modified problems in which the conditional premise was indirectly questioned by the third premise was enough to suggest a mistrustful attitude while solving the last standard argument. The representation of the task was modified from the first to the last standard argument. On the last one, fewer participants endorsed the conclusion than on the first one, which shows that some learnt that it was permissible to question the conditional. This suggests how easy it is to shift from one mode of treatment of the conditional, a trustful mode where it is not questioned, to another mode, a mistrustful one, where it is questioned through the sufficiency of its antecedent.

This result has a methodological consequence for studies of standard MT. Indeed the experiments described above by Cummins (who used causal rules) and Thompson (who used causal, definition and social rules) showed that endorsement of the conclusion of standard arguments depends on the availability of CNCs. Participants in these experiments necessarily had a representation of the task such that they spontaneously made use of their general knowledge of CNCs. The present results obtained with another manipulation (a suggestion induced by the sequence of problems) and other types of conditionals (means-end, remedial) as well as a causal, concur to support the idea that performance on MT using thematic material cannot be assessed without a control of the participants' perception of the extent to which they may apply their knowledge of CNCs to question the truth of the major conditional premise.

GENERAL DISCUSSION

Possible limitations of the study

The generality and robustness of the present results need to be examined. Generalisation across contexts has been considered in the discussion of the first experiment. Generalisation across arguments is another question. In the experiments, the number of problems was limited by time constraints. For purpose of generalisation of the results, we preferred to present several types of conditionals in one logical argument rather than the opposite because, in the studies that have been reviewed, both types of argument (MT and MP) were affected in the same manner by the various manipulations whereas in few of these studies, if any, was the type of conditional rule characterised and varied. Since the phenomena of interest, viz. the depressed rate of endorsement of the conclusion and the expression of uncertainty about the conclusion, are essentially and consistently equivalent for MT and MP across studies, there is no serious reason to doubt the generalisation of the

present results from MT to MP. A more relevant and uncertain problem regards the generalisation to more complex arguments. To the best of our knowledge, this has not yet been addressed and it is an open question for future research.

As far as robustness is concerned, two experiments followed the first one. In experiment 3 the modifications were minimal: one context and one condition were removed and the sample came from the same population; the overall pattern across contexts, as well as the individual patterns very closely replicated those observed in the first experiment. In experiment 2, one context and another condition were removed but the population was quite different; still, the results showed the same predicted pattern as the other two experiments. We can therefore conclude that, for MT arguments, the results are robust within the limits of the kind of contexts used.

Conditionals, CNCs and pragmatics

The role of Complementary Necessary Conditions was anticipated by a few authors in the past, witness this statement by Ramsey (1931): "In general we can say with Mill that 'if P then Q' means that Q is inferrable from P, that is, of course, from P *together with certain facts and laws not stated but in some way indicated by the context*" (emphasis added). Similarly, Goodman (1947) in his discussion of counterfactual conditionals stated that "the assertion that a connection holds is made on the presumption that certain circumstances not stated in the antecedent obtain" (p. 116). The status of unstated information indicated by the context could not be expounded at that time for lack of a well developed pragmatic theory.

Nowadays, this can easily be done. On the basis of a theory such as relevance theory (Sperber and Wilson, 1986), the assumption of satisfaction of CNCs (Ramsey's unstated facts and laws) can be characterised as an epistemically based implicature, which can be cancelled by further information. In effect, according to the principle of relevance, in uttering the conditional sentence, the speaker guarantees that the utterance is worth paying attention to. In the frame of a deductive argument, this amounts to a guarantee that an inference can be made, which requires that the CNCs be satisfied. Conditionals are typically uttered with an implicit *ceteris paribus* assumption to the effect that the normal conditions of the world (the satisfaction of the CNCs that belong to shared knowledge) hold. Should further information deny or just raise doubt on this assumption, the conditional premise no longer conveys a sufficient condition.

It is beyond the scope of this paper to give a detailed typology of CNCs; nevertheless, a few of their important features can be considered briefly. First, CNCs vary in the degree to which they are necessary for the consequent of the conditional to be true. Some are *sine qua non* conditions, whereas others are less

indispensable in the sense that their intervention does not affect the consequent in an all-or-nothing manner; rather, they render the consequent more or less likely to be true.

Next, there are two kinds of CNCs from the viewpoint of their polarity. Some of them are such that the consequent of the conditional cannot be the case (or is less likely) in their absence, whereas others are such that the consequent cannot be the case (or is less likely) in their presence. They are called enablers and disablers in the domain of causality, respectively.

Finally, not all CNCs are equally available. Some, limited in number, have high availability, whereas others have low availability but are virtually unlimited in number. It seems that a psychological criterion can help distinguish between the two. Upon being asked what is necessary for a match to light, most people would presumably agree on a few conditions easily retrieved from memory (e.g. , it should be dry, it should not break). If pressed for more conditions, different people would presumably produce different answers (e.g. , the match was not struck on the surface of the moon, it was not struck by a skin diver below surface) after some relatively long abductive process that results in preconditions rather than in conditions proper.

One of the main sources of possible disagreement over conditional statements is that the hearer may have independent reasons (different sources of information, or new evidence) to doubt the satisfaction of a CNC. This situation was exploited in George's (1995) experiments reviewed above. But such disagreement in no way contradicts the assumption of satisfaction of the CNC that concerns the attribution of an epistemic state to the speaker by the hearer: the speaker may be wrong however sincere they are.

In the experiments that have been reviewed, the implicature is cancelled or questioned in different ways: an additional premise can explicitly deny the CNC (e. g. Hilton & al. and the present experiments) or explicitly bring uncertainty about it (e. g. Stevenson & Over); or the CNC comes as the antecedent of another conditional premise (e. g. Byrne), which generates an epistemic implicature of uncertainty with respect to the CNC.

In other experiments, no additional premise is presented (Cummins & al. , 1995; Thompson, 1994, 1995) and this raises an interesting question: how can the satisfaction of the CNC be questioned? Clearly, the doubt on the CNC is generated by the reasoner herself/himself. But why should participants in this type of experiment behave differently to those of numerous experiments on conditional reasoning who endorse the conclusion of all MP arguments more than 95 percent of the time? Or, in other words, why are participants sensitive to the credibility of the

conditional premise in this type of experiment and not in more standard experiments?

At this point, the notion of task representation becomes crucial; it is linked to the reasoner's interaction with the experimenter. In all testing situations, whether in a school or an experimental setting, there is an element of pretense at work: the relevance of the response for the questioner does not lie in its informational content, but in its correctness, and this is shared knowledge with the participant. It is therefore essential for the latter to figure out what type of skill the experimenter is looking for (this helps eliminate some ways of responding, sometimes based on the judgment that 'it cannot be that easy'); this is the main determinant of the representation of the task. In brief, a response is worth giving when it is judged relevant enough to satisfy the expectation of relevance attributed by the participant to the experimenter. From this point of view, in the standard experiments on conditional reasoning, the instructions and the format of response make it clear that the experimenter is interested in the participant's deriving sure conclusions. Participants are almost always instructed that the premise must be considered as true, which is easy to follow if, as is generally the case, the premise is not controversial, that is, there are not too many available CNCs and their satisfaction is plausible enough. In that case, the assumption of satisfaction of the CNCs is maintained and the participant is involved in standard deductive reasoning.

Now, in Cummins' and Thompson's experiments the design was comparative, that is, participants received conditionals that had a great variability with respect to number and availability of CNCs and to the likelihood of their satisfaction. In such a situation, the task becomes fairly transparent: participants realise that they are invited to exploit and exhibit their sensitiveness to the differences in the credibility of the conditionals. They are licensed to cancel or not cancel the implicature. Since the guarantee from the speaker/experimenter is lifted, the participants are free to question the satisfaction of the CNC and they will do so based on world knowledge. The rate of endorsement of the conclusion decreased as the number of available CNCs increased because the more there are CNCs the lower the likelihood that they are all satisfied.

The third experiment presented in the present paper concerns the kind of situation where there is no additional premise. Participants had to solve a target standard MT after the solution of a first standard control MT and two non standard MT in which the third premise stated the non satisfaction of a CNC. A significant proportion shifted from the endorsement of the conclusion on the control MT to the non endorsement of the conclusion on the target MT. In the light of the foregoing discussion, the interpretation of this result is that the solution of the two non

standard MT was sufficient to suggest a change in the representation of the task by which it was permitted to doubt the satisfaction of the CNC.

Notice that even in the experiments that used a third premise to raise doubt on the satisfaction of the CNC (Byrne, 1989; Chan and Chua, 1994; Stevenson and Over, 1995), a sizeable proportion of participants (over one third) still endorsed the conclusion, which demonstrates that the decision to cancel the implicature that always lies in the reasoner's hands may not be so easy to make. The task representation is not clear to all because there are conflicting demands: on the one hand, assume the truth of the premises as required by the instructions, and on the other hand, doubt the truth of the conditional as suggested by the third premise.

A recent study by Vadeboncoeur and Markovits (1999) is highly relevant to this point. The authors presented MP and MT arguments in four different experimental conditions. The *Simple condition* was a kind of standard presentation in which the conditional premise (e. g. *If a match is struck then it lights*) was introduced by "*suppose it is true that . . .*" In the other three conditions, before the presentation of the argument, participants were told that even though it might not always be the case in everyday life, it was very important to suppose that the conditional premise was true. In the *Generation condition*, participants were required to write down one reason why the premise might not be true. In the *Explicit condition* such a reason was provided (e. g. the match is wet). In the *Logical condition*, no request or suggestion with regard to disabling conditions was made. In all four conditions the conclusion was to be evaluated by *it is certain that. . .* or "*we cannot know with certainty whether or not. . .*" It was observed that the rate of endorsement of the conclusion was the highest for the *Logical* group, the lowest for the *Simple* group, and intermediate for the *Generation* and *Explicit* groups. These results will be interpreted from the point of view of the present approach, which they illustrate well.

The difference between the *Simple* and the *Logical* groups lies in the instructions; by stressing that the premise should be considered as always true in the latter, the assumption of satisfaction of the CNC was made explicit, so leading to a higher frequency of *certain* conclusions than in the former. The other two situations, each in their own way, suggest or provide reasons to cancel the implicature of satisfaction of the CNC, following which an attempt at restoring full belief in the premise is made by asking participants to suppose that the statement is always true: the two messages are plainly contradictory, and depending on whether participants give more weight to the first or to the second, they will consider the premise as less certain or more certain. In other words, in between the *Simple* condition (no emphasis on the truth of the conditional, low endorsement of the conclusion) and the

Logical condition (emphasis, high endorsement), there are the other two, both of which are characterised by an emphasis on the truth of the conditional countermanded by the mention of the disabling conditions, hence an intermediate rate of endorsement.

The authors' aim was to demonstrate the role of disabling conditions retrieved in long term memory on causal conditional reasoning, a point in agreement with the present paper's position. There is, however, a point of disagreement. The authors considered a pragmatic explanation for the lower endorsement rate in the *Generation* and the *Explicit* conditions than in the *Logical* condition and rejected it. In their pragmatic explanation it is hypothesised that participants assume that the disabling factors proposed or suggested are relevant to the task, so contributing to less endorsement of the conclusion. The authors argue that if this was the case, there would be more endorsement in situations like the *Simple* condition in which the possibility of disabling conditions is less evident, which is contrary to what was observed. But, of course, this argument should apply *ceteris paribus*: it overlooks the heavily stressed instructions to accept the premise as true given to the *Logical* group but not to the *Simple* group, which goes counter to any effect of the disabling conditions. Therefore, it seems that the authors failed in their refutation. In brief, from the present viewpoint the exploitation of world knowledge and pragmatic-based interpretational phenomena are complementary (to such an extent that the latter could not operate without the former) whereas for the authors retrieval of information in long term memory and pragmatic factors (at least those alluded to, which seem elementary and are left unspecified) are mutually exclusive.

In summary, one condition of major importance to believe or doubt a conditional statement lies in the necessary conditions that are complementary to the antecedent⁹. The assumption of satisfaction of these conditions comes as an implicature of the conditional. When, as is the case in experimental manipulations that have been reviewed, such a complementary condition enters the context that has been set up by the conditional sentence (as the antecedent of a conditional, or as an explicit or an implicit denial), the implicature may be cancelled and the conditional becomes uncertain. The foregoing is an interpretational stage that delivers its output to the deductive component proper, whatever it is. The uncertainty will propagate to the conclusion, a process about which little is known except the investigations by Shultz, Zelazo, and Engelberg (1989) and more recently by George (1999)¹⁰.

⁹ It is not claimed that this is the only condition: in case the antecedent is perceived as irrelevant to the consequent, the statement has right away low credibility.

¹⁰ Although the present study has focused on deductively valid conditional arguments, the approach taken can generalise to the two well known fallacies of affirming the consequent and

Conditional probability and CNCs.

A classical thesis among linguists and philosophers is that belief in an indicative conditional can be measured by the conditional belief of the consequent on the antecedent. This raises the interesting question of how the present claim about the role of CNCs is related to this view. The simplest, and also strongest form of the conditional probability thesis is the plain equality, $\text{bel}(\text{if } A, C) = \text{bel } C/A$ as proposed by Adams (1975). Now, assuming that (together with more technical points that we need not consider) beliefs and conditional beliefs are subjective probabilities that follow the standard probability calculus, and that the left hand side of the equation represents probability of truth, it has been demonstrated by Lewis (1986) that this equation cannot be correct. In agreement with Lewis, many authors claim that $\text{bel } C/A$ is not a measure of $\text{bel}(\text{if } A, C)$ but a measure of the assertability of *if* A, C . There is an essential difference between the two. Whereas the former concerns truth conditions, the latter concerns the meaning of the conditional. Jackson (1987) explicates the difference by way of a useful comparison that concerns *and* and *but*. *But* has the same truth conditions and probability of truth as *and*; however, it has assertability conditions (assertibility, in his terminology) whose mastery can be exhibited by speakers' ability to say when it is right to use *but* rather than *and*: it conveys implicatures in addition to truth conditions. Similarly, the assertability conditions of *if* differ from its truth conditions. Asserting a conditional serves various pragmatic purposes, the most typical of which is to prepare the hearer for the development of a Modus Ponens. A high value for $\text{bel } C/A$ indicates the extent to which the conditional is exploitable to achieve such goals. Even if, from a pragmatic viewpoint, this analysis based on a Gricean approach is not fully developed, it will suffice for our current purposes. We are going to try to relate the theoretical approach of the present paper to the equation proposed by Jackson and Lewis in terms of assertability, namely: $\text{ass}(\text{if } A, C) = \text{bel } C/A$.

Notice first that Lewis argues that ordinary speakers' intuitions concern assertability, not probability of truth. If this is agreed, credibility of conditionals as

denying the antecedent. It has been argued earlier that the suppression of a valid conclusion results from the cancellation of an implicature - the one that guarantees the satisfaction of a CNC. Similarly, the suppression of an invalid conclusion (and consequently an improvement in performance) may result from the cancellation of a kind of implicature often called, after Geis and Zwicky (1971) an "invited inference" (whose effect is to provide a biconditional interpretation of the major premise). In fact, in a number of studies reviewed above, a counterpart of the manipulation on valid arguments was carried out on invalid arguments by adding a third premise whose aim was to provide an alternative antecedent to the major premise, so blocking the invited inference. Indeed, the effect of this manipulation was to diminish the fallacies. For a more comprehensive treatment of this phenomenon, see Politzer (in press).

considered in the three experiments above is assertability. Second, as Edgington (1995) argues, conditional belief might well be psychologically primitive; that is, it is psychologically implausible that people derive it by a calculation following its mathematical expression, $\text{bel } C/A = \text{bel } A \& C / \text{bel } A$, if only because people may have a value for $\text{bel } A/C$ without having a value for $\text{bel } A \& C$ or for $\text{bel } A$. In fact, even theorists such as Ramsey (1931) took conditional belief as a primitive from which belief in a conjunction is derived.

What is the origin of conditional belief? We are not going to answer this formidable question. We only wish to suggest that the consideration of a CNC is a factor that affects $\text{bel } C/A$. In order to show this, we will make an elementary calculation that will subsequently be illustrated in a diagram.

We will use the notation $A \rightarrow C$ for the conditional *if A, C*. When it is asserted with a *ceteris paribus* assumption to the effect that all relevant CNCs are satisfied, it can be rewritten as $A \& \{N\} \rightarrow C$ (where N represents the conjunction of the CNCs and braces indicate that N is implicitly satisfied). The assertability of the conditional is: $\text{bel } A \& \{N\} \& C / \text{bel } A \& \{N\}$. Since N is a necessary condition for C, N includes C and the ratio reduces to: $\text{bel } A \& C / \text{bel } A \& \{N\}$ (1). But when a CNC (call it N_1) is no longer satisfied, the assertability of the conditional $A \& N_1 \rightarrow C$ is: $\text{bel } A \& N_1 \& C / \text{bel } A \& N_1$, which again reduces to: $\text{bel } A \& C / \text{bel } A \& N_1$ (2) for the same reason as earlier. Comparing the denominators of (1) and (2) we can see that $\text{bel } A \& \{N\} < \text{bel } A \& N_1$ because N_1 includes $\{N\}$ and therefore: $\text{bel } A \& N_1 \& C / \text{bel } A \& N_1 < \text{bel } A \& C / \text{bel } A \& \{N\}$, which means that the assertability has decreased. In brief, the introduction of a CNC into the context (by cancelling the assumption of its satisfaction) decreases the assertability of the conditional. It is remarkable that the loss in assertability of the conditional sentence (understood as an intuitive pre-theoretical notion) that speakers experience upon hearing that a CNC hitherto assumed is not satisfied has an exact counterpart in the loss in assertability (understood as a theoretical construct) just demonstrated.

This demonstration can be interpreted in a diagram. We will assume, for the sake of simplicity, that there are only two CNCs, N_1 and N_2 (see Figure 1).

Insert Figure 1 about here

The conditional $A \rightarrow C$ is all the more assertable as the part of A that is common with C (i. e. , $A \& C$) is larger (with respect to the whole of A); or equivalently, $A \rightarrow C$ is all the more assertable as the part of A that lies outside C (the left "crescent") is smaller (again with respect to the whole of A). A property of CNCs is that they cut

out the left part of the "crescent". The more numerous the CNCs that are satisfied, the more reduced the part of A outside C that lies inside the intersection of the CNCs. When all relevant CNCs are satisfied, this part reaches a minimum: it is represented in Figure 1 by the dark area¹¹. CNCs and their intersection are implicit; they are represented by $\{N\}$. The assertability of the conditional $A \rightarrow C$ is represented by the ratio of the area $A \& C$ to the part of A that lies within $\{N\}$, i.e. $A \& C$ plus the dark area. When one of the CNCs (say N_1) is uncertain, it enters the context, stops being implicit and some part of A, namely the horizontally hatched area, is incorporated in $\{N\}$ (in fact, $\{N\}$ is extended to N_2). The conditional is $A \& N_1 \rightarrow C$; its assertability is the ratio of $A \& C$ divided by $A \& C$ plus the dark and horizontally hatched areas: it is diminished because the former denominator is increased by the horizontally hatched area. Notice that this view can also explain the concept of levels of sufficiency of a conditional in terms of conditional belief: the more sufficient the antecedent, the more numerous the (satisfied) CNCs, that is, the intersecting circles around C; and the more numerous the CNCs, the more reduced the part of A (the crescent) that lies outside C, that is, the smaller the uncertainty¹².

In brief, the concept of CNC together with the notion that they may or may not be satisfied appears to be compatible with (and possibly explainable by) theories that equate the assertability of a conditional with conditional belief. But a word of caution is in order. In this discussion, it has been tacitly assumed, like in most of the literature, that belief obeys the standard probability calculus. However, we do not wish to commit ourselves to such a strong hypothesis. We do not rule out the hypothesis that belief might be more adequately represented in a formalism using probability intervals such as the Dempster-Shafer theory (Shafer, 1976) or possibility theory (Dubois & Prade, 1988). But the question of the compatibility of our approach with such alternatives is beyond the scope of the present paper.

¹¹ In representing an area inside $A \& \{N\}$ but outside C, we allow for the possibility that the satisfaction of CNCs does not exhaust the source of credibility in $A \rightarrow C$: As noticed earlier (note 9), the relatedness of A and C is in itself a reason to believe or disbelieve the conditional. But the same demonstration could have been made without that area, *mutatis mutandis*.

¹² Similarly, the notion that some CNCs are more important than others can be accommodated within this framework: in that case, CNCs are not strictly necessary and the less important (call it N_1) covers a smaller part of C than the more important one (call it N_2). We have: $\text{bel } A \& N_2 \rightarrow C > \text{bel } A \& N_1 \rightarrow C$. With a simple additional hypothesis, it can be shown that when $\text{not-}N_1$ and $\text{not-}N_2$, respectively, are added as premises, $\text{bel } A \& \text{not-}N_2 \rightarrow C < \text{bel } A \& \text{not-}N_1 \rightarrow C$. This provides a formal justification for the manipulation underlying the first two experiments: the more important the CNC whose satisfaction is denied, the more the credibility of the associated conditional decreases.

The suppression effect

Byrne proposes that the suppression effect rests on the explicit representation of counterexamples to a conclusion. There are several important points of agreement between the counterexample approach and the uncertainty approach, but also points of disagreement. We consider both in turn.

First, we concur with Byrne (1991) that "in daily mental life there are always background conditions necessary for an outcome that can be called into question" (p. 77): CNCs are just such conditions for the consequent of a conditional to occur. Second, and as importantly because it is at the root of our pragmatic approach, we concur with Byrne, Espino, and Santamaría (1999, henceforth BES) that the interpretative component is essential to explain the suppression effect. Third, we fully agree with Byrne, Espino, and Santamaría (2000) that "the suppression of inferences arises because the provision of extra conditionals with additional or alternative antecedents changes the interpretation that reasoners reach of conditionals" (p. 105) but, as we will see shortly, we disagree on what the interpretation is changed into, and how it is produced. Finally, we are in agreement with BES (p. 351 and table p. 352) about the joint representation of the two conditionals *If A, C* and *If N, C* as a new conditional whose antecedent is the conjunction of their two antecedents, *A & N*, but again we disagree on how people reach it¹³.

In short, we disagree in that: (1) BES think that the integration of the conditional premises results from a process of construction of mental models based on compositional rules; we think that it results from a pragmatic process by which the additional premise *If N, C* conveys an epistemic implicature to the effect that N might not be satisfied, from which it follows that N becomes explicit in the antecedent, turning the major premise into *If A & N, C*. (2) BES think that the major premise is turned into a reversed conditional; we think that it is a direct conditional whose degree of belief has been lowered as a consequence of the first point above. (3) BES think that the conclusion is suppressed when people consider the situation in which A is the case but C is not the case because N does not obtain (counterexample availability); we think that the conclusion is endorsed with a degree of belief inherited from the premises.

¹³ In deep structure, the antecedent of this conditional is the conjunction of A and N, and the consequent is C. The theorist in her/his metalanguage as well as participants in experiments have to provide a surface structure. There are several possible expressions for this: one is *if A and if N, then C* (a formulation chosen by Politzer & Braine, 1991); another one is *if A and N, C*, as used by a majority of participants in the experiments of Byrne & Johnson-laird (1992). These conventional surface structure renderings should not obscure the problem, which concerns the deep structure on which there is agreement.

Is there a means of testing between the two approaches? This was the aim of BES's first experiment in which participants were asked to produce their own conclusion. The conclusions that were not endorsements of *C* (for MP) or of *not-A* (for MT) consisted of about one half of answers to the effect that there was not enough information to conclude (pooling together MP and MT), about 30% of answers reproducing the major premise *if A, C* (for MP) or expressing *not-A* or *not-N*, or else *not-A and not-N* (for MT) whereas hardly more than 2% of the answers were of the type *can-C/might-C*. (A study by Dieussaert, Schaeken, Schroyens, and d'Ydewalle (2000) yielded similar categories of answers). BES think that their results go against the uncertainty approach because, in their opinion, our approach should predict only the latter kind of answer. We think that their results support our point of view because over 90% of these answers are, to various degrees, means of expressing doubt about the conclusion. BES think that their results are consistent with their approach; we think that they are at best inconclusive in this respect because their predictions ("participants should tend to produce conclusions that refer to the additional condition that has not been affirmed or denied in the argument", p. 354) are too unprecise and also they did not give any account of the most frequent "not enough information" answer.

There were three other experiments in BES's paper. The last one included alternative conditionals and is not relevant to the present discussion. The second experiment aimed at testing the hypothesis of conversion of the additional premise formulated by Politzer and Braine (1991; note that this hypothesis is completely independent of the uncertainty hypothesis that was also outlined separately in the same paper). The major premise, or the additional premise, or both were expanded (in order to block the converse) like in the following example: (p1) *If Alicia met her friend then she went to the cinema*; (p2, expanded) *If she had money for a cinema-ticket then she went to the cinema, but if she went to the cinema she may or may not have had money for a cinema-ticket*; (p3) *Alicia met her friend*. In all three expanded conditions the suppression occurred. Indeed, this counts against the conversion hypothesis. Considered now from the uncertainty point of view, the result is immediately explainable. In effect, in the two conditions where the expansion concerns the additional premise (like the condition just presented) it is explicitly stated that the CNC may or maynot be the case. (In the other condition where the expansion concerns the major premise the suppression occurs because the additional premise is unchanged and plays its usual role).

The third experiment aimed at testing a prediction derived from mental models theory: a biconditional formulation of the additional premise should result in more suppression than the conditional formulation. The manipulation consisted in turning

either the major premise, or the additional premise, or both into explicit biconditionals (using *if and only if*). More suppression was observed in the two conditions in which the additional premise was a biconditional than in the conditions in which it was a conditional. BES think that this result supports the mental model theory, a claim that we do not dispute. We think that it supports the uncertainty approach as well. In effect, the biconditional makes it explicit that the CNC is a necessary condition, so that it suggests more strongly the epistemic implicature to the effect that the CNC may not be satisfied.

In fact, the joint results of two recent experiments carried out by Neth and Beller (1999) strongly support the uncertainty hypothesis. In one experiment, a control group was presented with various conditional relations *If P then Q* and an experimental group was presented with the same sentences plus a statement saying that no information about an appropriate CNC was available. Both groups had to rate their degree of confidence in *Q* provided that *P*. A significant decrease in confidence (13%) was observed for the experimental group: the sheer mention of a CNC that raised doubt about the satisfaction of the consequent affected belief in the consequent. But what about participants' assessment of the conditional relation? In another experiment participants were presented with *P*, *Q*, and asked, on the basis of their knowledge of the domain, to decide which of the following expressions was appropriate to express the relation between *P* and *Q*: (1) *If P then Q*; (2) *If Q then P*; (3) both of these; (4) none of these. As in the first experiment, participants received either an additional statement saying that "no information about [the CNC] is available" or no such statement (control). Limiting the results to relations independently judged as sufficient and non necessary, 97% of the participants in the control group chose *If P then Q* as the most appropriate expression of the relation; in contrast, only 50% did so in the experimental group, whereas 40% chose option (4). This means that for 40% of the participants the doubt on the CNC aroused by the reference made to it resulted in their dismissing the *If then* relation between *P* and *Q*, even though the same participants (the design was within-subjects) nearly unanimously found the *If then* relation appropriate when there was no mention of the CNC¹⁴.

To conclude this section, we summarise why we prefer the uncertainty approach to the suppression effect rather than the counterexample approach. We state first, as an epistemological prerequisite, why we are in a position to explain it. It has been shown earlier that when the credibility of a premise is manipulated in

¹⁴ These percentages are typical of the studies reviewed: belief in the conditional was maintained by one half of the participants; it was given up by slightly less (40%).

various ways as reviewed (in particular by affecting a CNC), belief in the conclusion varies in a systematic and predictable manner, namely as an increasing function of credibility. This does not mean that any decrease in the rate of endorsement of the conclusion (or in its credibility) following the manipulation of a premise necessarily implies lowered credibility of a premise. It is only suggestive of that. But if, in addition, one is able to show independently (here based on pragmatic theory) that the manipulation did just that, then one possesses a likely explanation of the phenomenon.

Theoretically, the uncertainty approach is based on very general, independent, pragmatic principles. It subsumes the suppression effect under a general class of situations of plausible reasoning: it concerns the particular case where belief in a conditional sentence is altered by contextual information. It takes a serious view on the fact that most of our everyday reasoning occurs with uncertain premises. Empirically, the review reported in this section shows that it has well resisted a few tests that could have refuted it (but of course, much more work is needed to confirm this latter point).

In contrast, it seems to us that there are two major difficulties with the counterexample account. First, a counterexample to a conclusion is an all-or-nothing concept by which the denial of the conclusion is delivered in its turn in an all-or-nothing manner. It cannot explain the expression of belief in the conclusion by degrees observed in experiments reviewed above, including the present ones. In a word, it is not equipped to account for one of the essential characteristics of the effect it aims to explain.

Second, and more fundamentally, it seems to us that it is essentially limited in scope: while the situation in which *the library does not stay open* is a clear counterexample to the MP conclusion *she will study late in the library*, for MT no such mechanism can operate: the situation in which *the library does not stay open* cannot count as a counterexample to the MT conclusion *she does not have textbooks to read*. We view this as an insuperable difficulty for this account.

Degrees of belief and theories of reasoning

The approach advocated here is driven by considerations that are mostly linguistic in nature: before considering the reasoning process proper, one should characterise the propositions that constitute its input. This has been our main concern so far, in terms of logical form and of degrees of belief. But what about the deductive process and the representation of belief?

It is just a plain fact that, in daily life, people draw inferences from uncertain premises. They are also capable of indicating a degree of belief in their conclusion

(which seems to be linked to the degree of belief in the premises in a systematic manner): this has been confirmed in studies reviewed above as well as in the present experiments. Explaining how this is possible is a serious challenge to all theories of reasoning. We do not see how to explain the transmission of degrees of belief from premises to conclusion in the framework of mental model theory: it seems to us that because propositions are broken into their atomic components and into their various models, any certainty value attached to a premise will be dissolved in the course of this process.

What explanation could be offered in the framework of mental logic? Mental logic has built-in devices that can, in principle, be exploited in order to accommodate belief representation and transmission.

First, consider the use of conditional sentences. They can be viewed as devices in charge of the processing of uncertain atomic propositions that they accommodate in their antecedent. Asserting a conditional $A \rightarrow C$ typically communicates that the speaker (1) believes that there is a link between A and C and that the degree of contingency of C on A is measured by $\text{bel } C/A$, and (2) typically (although not necessarily) is uncertain about A.

In addition, it is understood that Modus Ponens is licensed by the assertion of the conditional. In general, MP delivers a lower bound for belief in C. There are several important particular cases. When A is fully believed, the conclusion C has credibility equal to that of the conditional. When the conditional is fully believed, the lower bound for $\text{bel } C$ is $\text{bel } A$: $\text{bel } C \geq \text{bel } A$. Finally, when both premises are fully believed we have a standard MP that delivers full belief in C. These are classic equalities or inequalities resulting from the conditional probability view: they may be stronger than we need.

Before going into more details, recall that in rule theory (Braine & O'Brien, 1998; Rips, 1994) the conditional is involved in two basic inference schemas: (1) elimination, by way of MP; (2) introduction, by way of conditionalisation (or "conditional proof"), that is, if C is derivable from a set of premises under the supposition that A, *if A then C* is a valid proposition. We suggest that the use of these two schemas is sufficient to deal with transmission of uncertainty. We have just considered the role of MP. Whereas MP conveys uncertainty to the conclusion (isolating the source of the uncertainty while eliminating the connective), the role of conditionalisation is to take charge of the uncertainty by the *introduction* of an antecedent together with the connective: Given P (a set of premises) and A uncertain, assuming A and deriving C allows one to assert *if A, C* where A is at its right place *qua* antecedent of a conditional. But there is more to say about conditionalisation: the very fact that a proposition is uncertain

suggests conditionalisation precisely because the rule system is defined for bi-valued propositions and in order to be processed the proposition has to be assumed. It follows that plausible reasoning is not hampered by the relative lack of availability of the schema for conditional proof for ordinary reasoners (in its strategic use) because its need is made salient by the processing of uncertain propositions¹⁵.

This general framework will now be applied to the kind of plausible reasoning that has been considered in this paper, and more specifically to the suppression effect and to the present experiments. We begin with MP and write two possible derivations of the response (1) and (2) as follows:

$\{N \&\} A \rightarrow C$		major premise (N in braces indicates an implicit CNC)
$N \rightarrow C$		additional premise
$N^\bullet \& A \rightarrow C$		N enters the context as an uncertain CNC
		and is made explicit in the major premise; (the black dot indicates an uncertain proposition)
A		minor premise
either	(1)	Stop derivation for lack of information on N. State "not enough information".
or	(2)	
	$[N^\bullet]$	conditionalisation
	$N^\bullet \& A$	conjunction
	C	(by MP to the three preceding lines)
	<hr/>	
	$N^\bullet \rightarrow C$	conclusion

These two conclusions are the ones that constitute around 90% of the answers obtained by BES and by Dieussaert et al. Of course, the case in which the conclusion is endorsed corresponds to the case (not written in the derivation above) where N is not taken into account (the assumption of satisfaction is not cancelled) and MP is directly applied to $A \rightarrow C$ and A, yielding C.

In the paradigms where belief in the conclusion is to be evaluated, N is reiterated and C follows, inheriting a degree of belief related to that of N, which is a function of $\text{bel } N$. (This could be specified on the basis of some hypotheses: suppose the general relation $\text{bel } C \geq \text{bel } N$ holds; then, if N is a strictly necessary condition, and if reasoners are aware of it, $\text{bel } N \geq \text{bel } C$ and therefore it follows that

¹⁵ The schema for conditionalisation is not valid in a probabilistic logic such as Adams' (1966), which entails the non validity of a number of important rules; but in this system, soundness is defined by a much stronger criterion than we need, and the related definition of uncertainty of a proposition A as $1 - p(A)$ differs from the one we are using, which is akin to entropy.

$\text{bel } C = \text{bel } N$).

We now turn to MT (for which the first three lines are the same as above):

$\{N \& \} A \rightarrow C$ major premise

$N \rightarrow C$

$N^\bullet \& A \rightarrow C$

not- C minor premise

either (1) Stop derivation for lack of information on N . State
"not enough information".

or (2) not $(N^\bullet \& A)$ (abbreviated derivation)

not- N^\bullet OR not- A conclusion¹⁶

Again, when N is not taken into account, a standard MT is applied to the first premise to yield or endorse *not-A*. A variant of this is an application of a standard MT to the first premise, and separately to the second one, yielding *not-A AND not-N*. Again, the three answers, *not enough information*, *not-A OR not-N*, *not-A AND not-N* cover around 90% of the observations made by the same authors in the production task.

In the tasks where belief in the conclusion *not-A* is to be evaluated, we assume that participants start from

not-N[•] OR not-A at which they have arrived spontaneously or after being prompted by the question. In saying earlier in the introduction that belief in A is disbelief in *not-A*, we anticipated a principle that can be stated more generally: all systems of belief must have some kind of truth conservation principle, the strongest of which is that of the probability calculus, $p(A) + p(\text{not-}A) = 1$. A weaker principle that does not assume complementarity and that is sufficient for our purposes is: *bel A is a decreasing function of bel not-A*; we assume a principle of this type to be universal. Having $A \text{ OR } B$ as true, and $B = \text{not-}A$, the greater the belief in A , the smaller the belief in B . Considering now the conclusion *not-N[•] OR not-A*, this means that the greater *bel not-N*, the smaller *bel not-A*, or equivalently the more people doubt the satisfaction of N , the less they believe the conclusion *not-A*.

We have thus suggested, within rule theory, an explanation for MT as well as for MP for the production and the evaluation of the response, that is the paradigms of the suppression effect and of plausible conditional reasoning, the former being, as has been claimed throughout this paper, a particular case of the latter. We wish to emphasise that in this section our aim has been limited to showing that, in principle, it is feasible to process uncertainty within rule theory. We have not engaged

¹⁶ The derivation to *not-N[•] OR not-A* follows from an application of a schema equivalent to one of De Morgan's laws. It belongs to Rips' rule system and its status is borderline in Braine's system.

ourselves in the formidable enterprise of describing a full-fledged theory of plausible reasoning, if only because we would need to specify beforehand the properties of belief from the point of view of measurement theory, then defining with precision some strict criterion for a valid plausible argument: this is for future research.

Acknowledgements

The authors are grateful to B. Walliser and L. Bonatti for many discussions and for their comments and suggestions on the manuscript.

References

- Adams, E. W. (1966). Probability and the logic of conditionals. In J. Hintikka & P. Suppes (Eds.), *Aspects of inductive logic* (pp. 265-316). Amsterdam: North-Holland.
- Adams, E. W. (1975). *The logic of conditionals*. Dordrecht: Reidel.
- Adams, E. W. , & Levine, H. P. (1975). On the uncertainties transmitted from premises to conclusions in deductive inferences. *Synthese*, 30, 429-460.
- Braine, M. D. S. , & O'Brien, D. P. (1998). *Mental Logic*. Mahwah: Lawrence Erlbaum Associates.
- Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, 31, 61-83.
- Byrne, R. M. J. (1991). Can valid inferences be suppressed? *Cognition*, 39, 71-78.
- Byrne, R. M. J. , Espino, O. , & Santamaría, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language*, 40, 347-373.
- Byrne, R. M. J. , Espino, O. , & Santamaría, C. (2000). Counterexample availability. In Schaeken, W. , De Vooght, G. , & d'Ydewalle, G. (Eds.), *Deductive reasoning and strategies* (pp. 97-119). London: Lawrence Erlbaum Associates.
- Byrne, R. M. J. , & Johnson-Laird, P. N. (1992). The spontaneous use of propositional connectives. *Quarterly Journal of Experimental Psychology*, 45A, 89-110.
- Chan, D. , & Chua, F. (1994). Suppression of valid inferences: Syntactic views, mental models, and relative salience. *Cognition*, 53, 217-238.
- Cummins, D. D. (1995). Naive theories and causal deduction. *Memory and Cognition*, 23, 646-658.
- Cummins, D. D. , Lubart, T. , Alksnis, O. , & Rist, R. (1991). Conditional reasoning and causation. *Memory and Cognition*, 19, 274-282.
- Dieussaert, K. , Schaeken, W. , De Neys, W. , & d'Ydewalle, G. (2000). Initial belief state as predictor of belief revision. *Current Psychology of Cognition*, 19, 274-282.
- Dieussaert, K. , Schaeken, W. , Schroyens, W. , & d'Ydewalle, G. (2000). Strategies during complex conditional inferences. *Thinking and Reasoning*, 6, 125-160.
- Dubois, D. , & Prade, H. (Eds.), (1988). *Possibility theory: An approach to computerized processing of uncertainty*. New York: Plenum Press.
- Edgington, D. (1995). On conditionals. *Mind*, 104, 235-329.
- Elio, R. (1997). What to believe when inferences are contradicted: The impact of knowledge type and inference rule. *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 211-216). Mahwah, N. J. : Lawrence Erlbaum.
- Elio, R. (1998). How to disbelieve $P \rightarrow Q$: Resolving contradictions. *Proceedings of the*

- Twentieth Annual Conference of the Cognitive Science Society* (pp. 315-320). Mahwah, N. J. : Lawrence Erlbaum.
- Elio, R. , & Pelletier, F. J. (1997). Belief change as propositional update. *Cognitive Science*, 21, 419-460.
- Evans, J. St. B. T. , & Twyman-Musgrove, J. (1998). Conditional reasoning with inducements and advice. *Cognition*, 69, B11-B16.
- Geis, M. L. , & Zwicky, A. M. (1971). On invited inferences. *Linguistic Inquiry*, 2, 561-566.
- George, C. (1995). The endorsement of the premises: Assumption-based or belief-based reasoning. *British Journal of Psychology*, 86, 93-111.
- George, C. (1997). Reasoning from uncertain premises. *Thinking and Reasoning*, 3, 161-189.
- George, C. (1999). Evaluation of the plausibility of a conclusion derivable from several arguments with uncertain premises. *Thinking and reasoning*, 5, 245-281.
- Goodman, N. (1947). The problem of counterfactual conditionals. *Journal of Philosophy*, 44, 113-128.
- Hilton, D. J. , Jaspars, J. M. F. , & Clarke, D. D. (1990). Pragmatic conditional reasoning: Context and content effects on the interpretation of causal assertions. *Journal of Pragmatics*, 14, 791-812.
- Jackson, F. (1987). *Conditionals*. Oxford: Basil Blackwell.
- Lewis, D. (1986). *Philosophical papers*. Vol 2. Oxford: Oxford University Press.
- Liu, I.-M. , Lo, K.-C. , & Wu, J.-T. (1996). A probabilistic interpretation of "If-Then". *Quarterly Journal of Experimental Psychology*, 49A, 828-844.
- Manktelow, K. I. , & Fairley, N. (2000). Superordinate principles in reasoning with causal and deontic conditionals. *Thinking and Reasoning*, 6, 41-65.
- Neth, H. , & Beller, S. (1999). How knowledge interferes with reasoning - Suppression effects by content and context. *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* (pp. 468-473). Mahwah, N. J. : Lawrence Erlbaum Associates.
- Newstead, S. E. , Ellis, M. C. , Evans, J. St.B. T. , & Dennis, I. (1997). Conditional reasoning with realistic material. *Thinking and Reasoning*, 3, 49-76.
- O'Brien, D. P. , Costa, G. , & Overton, W. F. (1986). Evaluations of causal and conditional hypotheses. *Quarterly Journal of Experimental Psychology*, 38A , 493-512.
- Politzer, G. (in press). Premise interpretation in conditional reasoning. In D. Hardman & L. Macchi (Eds.), *Reasoning and decision making: A handbook*. Wiley.
- Politzer, G. , & Braine, M. D. S. (1991). Responses to inconsistent premisses cannot count as suppression of valid inferences. *Cognition*, 38, 103-108.
- Politzer, G. , & Carles, L. (2001). Belief revision and uncertain reasoning. *Thinking*

- and Reasoning*, 7, 217-234.
- Pollock, J. L. (1987). Defeasible reasoning. *Cognitive Science*, 11, 481-518.
- Ramsey, F. P. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays* (pp. 156-198). London: Routledge and Kegan Paul.
- Rescher, N. (1976). *Plausible reasoning*. Amsterdam: Van Gorcum.
- Rips, L. J. (1994). *The psychology of proof*. Cambridge: MIT Press.
- Shafer, G. (1976). *A mathematical theory of evidence*. Princeton: Princeton University Press.
- Shultz, T. R. , Zelazo, P. D. & Engelberg, D. J. (1989). Managing uncertainty in rule-based reasoning. *Eleventh Annual Conference of the Cognitive Science Society* (pp. 227-234). Hillsdale, N.J. : Lawrence Erlbaum.
- Sperber, D. , & Wilson, D. (1986). *Relevance: Communication and cognition*. London: Blackwell.
- Stevenson, R. J. , & Over, D. E. (1995). Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology*, 48A, 613-643.
- Suppes, P. (1966). Probabilistic inference and the concept of total evidence. In J. Hintikka & P. Suppes (Eds.), *Aspects of inductive logic* (pp. 49-65). Amsterdam: North-Holland.
- Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory and Cognition*, 22, 742-758.
- Thompson, V. A. (1995). Conditional reasoning: The necessary and sufficient conditions. *Canadian Journal of Experimental Psychology*, 49, 1-60.
- Vadeboncoeur, I. , & Markovits, H. (1999). The effect of instruction and information retrieval on accepting the premises in a conditional reasoning task. *Thinking and Reasoning*, 5, 97-113.

Appendix

Experiment 1

The five contexts (translated from French) in each of the five conditions. The complete set of premises and the contextual sentences are given in full for one context only (Headache). For the other four contexts, only the additional premise (Ad) is given. (M) stands for major premise, (m) for minor premise, (C) for conclusion.

CONTEXT <HEADACHE>

While talking with the wife of your friend Paul, you get to know that he has a headache at work nearly every day. You know that:

1) Standard condition:

(M) If Paul has a headache then he takes aspirin

Paul arrives and he says that

(m) He has not taken aspirin

What do you think of the following conclusion:

(C) Paul did not have a headache

2) High credibility (necessary condition satisfied):

(M) If Paul has a headache then he takes aspirin

As the conversation goes on, you get to know that

(Ad) Paul always has water with him

Paul arrives and he says that

(m) He has not taken aspirin

What do you think of the following conclusion:

(C) Paul did not have a headache

3) Low credibility (necessary condition unsatisfied):

(M) If Paul has a headache then he takes aspirin

As the conversation goes on, you get to know that

(Ad) Paul does not always have water with him

Paul arrives and he says that

(m) He has not taken aspirin

What do you think of the following conclusion:

(C) Paul did not have a headache

4) Very low credibility (strict necessary condition unsatisfied):

(M) If Paul has a headache then he takes aspirin

As the conversation goes on, you learn that

(Ad) The doctor has just forbidden him to take aspirin

Paul arrives and he says that

(m) He has not taken aspirin

What do you think of the following conclusion:

(C) Paul did not have a headache

5) *Explicit-normality condition:*

(M) If Paul has a headache then he takes aspirin

As the conversation goes on, you learn that

(Ad) All the conditions for him to take aspirin were satisfied

Paul arrives and he says that

(m) He has not taken aspirin

What do you think of the following conclusion:

(C) Paul did not have a headache

CONTEXT <BAKER'S>

1) Standard condition:

(M) If Helen goes to the baker's then she brings back bread

(m) Helen has not brought back bread

(C) Helen has not been to the baker's

2) High credibility:

(Ad) There is a lot of bread to day

3) Low credibility :

(Ad) There is not much bread today

4) Very low credibility:

(Ad) The baker's shop is exceptionally closed today

5) *Explicit-normality condition:*

(Ad) All the conditions for her to bring back bread were satisfied

CONTEXT <FISHING>

1) Standard condition:

(M) If Michael goes fishing then he has a fish meal

(m) Michael did not have a fish meal

(C) Michael did not go fishing

2) High credibility:

(A) There are many fish in the river currently

3) Low credibility:

(Ad) There are not many fish in the river currently

4) Very low credibility:

(Ad) Michael cannot stand eating fish any longer

5) Explicit-normality condition:

(Ad) All the conditions for him to have a fish meal were satisfied

CONTEXT <ALARM>

1) Standard condition:

(M) If somebody touches an object on display then the alarm is set off

(m) The alarm was not set off

(C) Nobody touched an object on display

2) High credibility:

(Ad) There was no problem with the equipment

3) Low credibility:

(Ad) There were some problems with the equipment

4) Very low credibility:

(Ad) The material was totally out of order

5) Explicit-normality condition:

(Ad) All the conditions for the alarm to be set off were satisfied

CONTEXT <TRAFFIC >

1) Standard condition:

(M) If the traffic is light then Mary takes her car

(m) Mary has not taken her car

(C) The traffic is not light

2) High credibility:

(Ad) Mary likes driving very much

3) Low credibility:

(Ad) Mary does not like driving

4) Very low credibility:

(Ad) Mary has been disqualified from driving

5) Explicit-normality condition:

(Ad) All the conditions for her to take her car were satisfied

Table 1.

Experiment 1. Distributions of responses (in percent) for various levels of credibility of the conditional. For each context N = 24 per row.

Context: <Headache>

Credibility of condi- tional	True	Evaluation of conclusion			
		Probably True	Indeter- minable	Probably False	False
High	50.0	50.0	00.0	00.0	00.0
Standard	41.7	45.8	12.5	00.0	00.0
Low	04.2	20.8	54.2	16.7	04.2
Very Low	00.0	04.2	25.0	50.0	20.8
Explicit- normality	45.8	16.7	04.2	12.5	20.8

Context: <Baker's>

Credibility of condi- tional	True	Evaluation of conclusion			
		Probably True	Indeter- minable	Probably False	False
High	70.8	20.8	04.2	00.0	04.2
Standard	58.3	29.2	08.3	04.2	00.0
Low	04.2	12.5	66.7	12.5	04.2
Very Low	04.2	00.0	45.8	25.0	25.0
Explicit- normality	54.2	29.2	08.3	00.0	08.3

Table 1. (continued)

Experiment 1. Distributions of responses (in percent) for various levels of credibility of the conditional. For each context N = 24 per row.

Context: <Fishing>

Credibility of condi- tional	True	Evaluation of conclusion			
		Probably True	Indeter- minable	Probably False	False
High	29.2	41.7	16.7	08.3	04.2
Standard	33.3	41.7	16.7	04.2	04.2
Low	04.2	04.2	58.3	25.0	08.3
Very Low	04.2	04.2	41.7	33.3	16.7
Explicit- normality	33.3	37.5	12.5	12.5	04.2

Context: <Alarm>

Credibility of condi- tional	True	Evaluation of conclusion			
		Probably True	Indeter- minable	Probably False	False
High	70.8	20.8	04.2	04.2	00.0
Standard	58.3	29.2	12.5	00.0	00.0
Low	04.2	25.0	45.8	12.5	12.5
Very Low	00.0	04.2	33.3	37.5	25.0
Explicit- normality	70.8	16.7	08.3	00.0	04.2

Table 1. (continued)

Experiment 1. Distributions of responses (in percent) for various levels of credibility of the conditional. For each context N = 24 per row.

Context: <Traffic>

Credibility of conditional	True	Evaluation of conclusion			
		Probably True	Indeterminable	Probably False	False
High	58.3	37.5	04.2	00.0	00.0
Standard	25.0	58.3	08.3	04.2	04.2
Low	04.2	08.3	66.7	12.5	08.3
Very Low	00.0	00.0	29.2	20.8	50.0
Explicit-normality	08.3	29.2	04.2	12.5	45.8

All contexts pooled (N = 120 per row)

Credibility of conditional	True	Evaluation of conclusion			
		Probably True	Indeterminable	Probably False	False
High	55.8	34.2	05.8	02.5	01.7
Standard	43.3	40.8	11.7	02.5	01.7
Low	04.2	14.2	58.3	15.8	07.5
Very Low	01.7	02.5	35.0	33.3	27.5
Explicit-normality	42.5	25.8	07.5	07.5	16.7

Table 2.

Experiment 2. Distributions of responses (in percent) for various levels of credibility of the conditional.

Credibility of conditional	True	Evaluation of conclusion			
		Probably True	Indeterminable	Probably False	False
High	43.5	28.2	16.1	09.7	02.4
Standard	28.0	42.7	18.7	04.0	06.7
Low	11.5	09.8	44.3	16.4	18.0

Table 3.

Experiment 3. Response choices (in percent) as a function of the position of the standard problem (first or last of the four) and as a function of the credibility of the conditional. For each context N = 30 per row.

Context: <Headache>

Credibility of conditional		Response				False
		True	Probably true	Indeter- minable	Probably false	
Standard	(First)	46.7	36.7	13.3	00.0	03.3
Standard	(Last)	30.0	46.7	20.0	03.3	00.0
Low		00.0	23.3	56.7	06.7	13.3
Very Low		03.3	06.7	33.3	40.0	16.7

Context: <Baker's>

Credibility of conditional		Response				False
		True	Probably true	Indeter- minable	Probably false	
Standard	(First)	60.0	33.3	06.7	00.0	00.0
Standard	(Last)	36.7	40.0	20.0	03.3	00.0
Low		03.3	13.3	70.0	10.0	03.3
Very Low		06.7	06.7	36.7	26.7	23.3

Context: <Fishing>

Credibility of conditional		Response				False
		True	Probably true	Indeter- minable	Probably false	
Standard	(First)	40.0	36.7	13.3	03.3	06.7
Standard	(Last)	23.3	43.3	30.0	03.3	00.0
Low		00.0	20.0	50.0	20.0	10.0
Very Low		03.3	06.7	40.0	30.0	20.0

Table 3 (continued).

Experiment 3. Response choices (in percent) as a function of the position of the standard problem (first or last of the four) and as a function of the credibility of the conditional. For each context N = 30 per row.

Context: <Alarm>

		Response				
		True	Probabl y true	Indeter- minable	Probabl y false	
Credib ility of cond itional						
Standard	(First)	63.3	26.7	10.0	00.0	00.0
Standard	(Last)	46.7	33.3	16.7	00.0	03.3
Low		06.7	20.0	53.3	13.3	06.7
Very Low		03.3	10.0	40.0	30.0	16.7

All contexts pooled (N = 120 per row).

Credibility of conditional		True	Probably true	Indeterminable	Probably false	False
Standard	(First)	52.5	33.3	10.8	00.8	02.5
Standard	(Last)	34.2	40.8	21.7	02.5	00.8
Low		02.5	19.1	57.5	12.7	08.3
Very Low		04.1	07.5	37.5	31.7	19.2

Table 4.

Experiment 3. Number of subjects who endorsed (*yes*) or did not endorse (*no*) the conclusion of the standard argument as a function of its position.

		<u>Endorsed in 1st position</u>		
endorsed in second position		yes	no	total
	no	30	49	79
	yes	33	8	41
	total	63	57	120

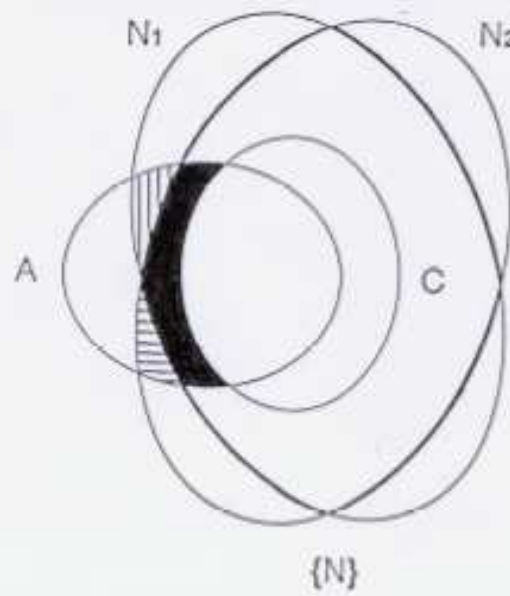


Figure 1. Conditional probability and complementary necessary conditions.

